Robert Reichelt, Tamara Rothmeier, Felix Grünberger, Sarah Willkomm, Astrid Bruckmann, Winfried Hausner and Dina Grohmann*

The archaeal Lsm protein from *Pyrococcus furiosus* binds co-transcriptionally to poly(U)-rich target RNAs

https://doi.org/10.1515/hsz-2023-0215
Received May 14, 2023; accepted August 22, 2023;
published online September 15, 2023

Abstract: Posttranscriptional processes in Bacteria include the association of small regulatory RNAs (sRNA) with a target mRNA. The sRNA/mRNA annealing process is often mediated by an RNA chaperone called Hfq. The functional role of bacterial and eukaryotic Lsm proteins is partially understood, whereas knowledge about archaeal Lsm proteins is scarce. Here, we used the genetically tractable archaeal hyperthermophile *Pyrococcus furiosus* to identify the protein interaction partners of the archaeal Sm-like proteins (PfuSmAP1) using mass spectrometry and performed a transcriptome-wide binding site analysis of PfuSmAP1. Most of the protein interaction partners we found are part of the RNA homoeostasis network in Archaea including ribosomal proteins, the exosome, RNA-modifying enzymes, but also RNA polymerase subunits, and transcription factors. We show that PfuSmAP1 preferentially binds messenger RNAs and antisense RNAs recognizing a gapped poly(U) sequence with high affinity. Furthermore, we found that SmAP1 co-transcriptionally associates with target RNAs. Our study reveals that in contrast to bacterial Hfq, PfuSmAP1 does not affect the transcriptional activity or the pausing behaviour of archaeal RNA polymerases. We propose that PfuSmAP1 recruits antisense RNAs to target mRNAs and thereby executes its putative regulatory function on the post-transcriptional level.

Keywords: Sm-like protein; SmAP; RNA-binding protein; Archaea; RNA polymerase; transcription

***Corresponding author: Dina Grohmann**, Institute of Microbiology & Archaea Centre, Single-Molecule Biochemistry Lab, University of Regensburg, D-93053 Regensburg, Germany; and Regensburg Center of Biochemistry (RCB), University of Regensburg, D-93053 Regensburg, Germany, E-mail: dina.grohmann@ur.de
Robert Reichelt, Tamara Rothmeier, Felix Grünberger, Sarah Willkomm and Winfried Hausner, Institute of Microbiology & Archaea Centre, Single-Molecule Biochemistry Lab, University of Regensburg, D-93053 Regensburg, Germany, E-mail: robert.reichelt@ur.de (R. Reichelt), Tamara.Rothmeier@stud.uni-regensburg.de (T. Rothmeier), Felix.Gruenberger@ur.de (F. Grünberger), sarah.willkomm@googlemail.com (S. Willkomm), Winfried.Hausner@ur.de (W. Hausner). https://orcid.org/0000-0003-3455-1495 (R. Reichelt)
Astrid Bruckmann, Institute of Biochemistry, Genetics and Microbiology (Biochemistry I), Protein Mass Spectrometry Laboratory, University of Regensburg, D-93053 Regensburg, Germany, E-mail: Astrid.Bruckmann@ur.de

1 Introduction

Members of the Sm and Sm-like (Lsm) protein family are conserved across all domains of life (Bacteria, Archaea and Eukarya) and are involved in cellular RNA metabolism (Mura et al. 2013; Wilusz and Wilusz 2005). Sm and Sm-like proteins form homomeric or heteromeric complexes resembling a doughnut-shaped ring (Kambach et al. 1999; Mura et al. 2001; Schumacher et al. 2002). They are characterized by an SM core fold that is constituted by two conserved Sm motif domains (Sm1 and Sm2). In contrast to the Sm core, the C- and N-terminal regions as well as the linker between the Sm1 and Sm2 motif are not conserved.

In Eukarya, at least 16 different Sm and Lsm proteins have been identified. These proteins typically form heteroheptameric Lsm1-7 complexes (Albrecht and Lengauer 2004; Wilusz and Wilusz 2005). Cytoplasmic Lsm1-7 complexes bind 3′UTRs of deadenylated mRNAs modulating mRNA decapping and decay (Tharun et al. 2000). Nuclear Sm and Lsm multimers are prominent members of the spliceosome (Zhan et al. 2018). The Lsm2-8 ring binds U-rich sequences in the 3′ UTR of the small nuclear (sn) RNA U6 ultimately supporting splicing and maturation of RNAs (Kufel et al. 2002). Beyond this, eukaryotic Sm and Lsm complexes are involved in several other RNA-mediated processes mainly acting as ribonucleoprotein (RNP) scaffolds (Mura et al. 2013).

In Bacteria, Lsm proteins are commonly referred to as host factor Q (Hfq) proteins, which are encoded in approximately 50 % of bacterial genomes (Franze de Fernandez et al. 1968; Valentin-Hansen et al. 2004). As RNA chaperones, Hfq proteins mediate interactions between small non-coding RNAs (ncRNAs) and messenger RNAs (mRNAs) and influence the folding landscape of RNAs even in a co-transcriptional

manner (Cai et al. 2022; Małecka and Woodson 2021; Rodgers et al. 2023). Thereby, they modulate translation either by masking or exposing the ribosomal binding site of the mRNA, which results in either translational repression or stimulation, degradation, or mRNA stabilization (Møller et al. 2002; Vogel and Luisi 2011). Hfq proteins form homohexameric complexes with distinct interaction surfaces for different RNA substrates (Mikulecky et al. 2004). They recognize RNAs with a poly(A) and poly(U) stretch, which are bound on the proximal and distal face of Hfq, respectively (Matera et al. 2007; Mura et al. 2003a,b). Beyond their role in the regulation of translation, under stress conditions Hfq has been shown to play a role in Rho-dependent transcription termination in *Escherichia coli*. It prevents premature binding of the termination factor Rho to the 5′ UTR of a subset of genes (Sedlyarova et al. 2016). This anti-termination mechanism requires the presence of small regulatory RNAs that are expressed under stress. Co-transcriptional binding of a small RNA and Hfq to the 5′ UTR stimulates transcription during the stationary phase transition (Kambara et al. 2018). Apart from this, *E. coli* Hfq is involved in ribosome assembly, assisting in the processing and folding of the 16S rRNA and the biogenesis of the 30S particle (Andrade et al. 2018).

Genes encoding archaeal Sm-like proteins (SmAPs) are found in all archaeal groups (Reichelt et al. 2018). Typically, at least one Lsm gene is detectable in archaeal genomes, which is often referred to as SmAP1 (Mura et al. 2013). In case of *Haloferax volcanii* (Hvo), SmAP1 protein forms homoheptameric complexes and binds U-rich RNAs *in vitro*. It also associates with small RNAs (sRNAs) and numerous protein complexes *in vivo* (Fischer et al. 2010). Deletion of the Sm1 motif in SmAP1 in *H. volcanii* resulted in major changes in the transcriptome and increased swarming activity (Maier et al. 2015). A second *smap* gene encoding a SmAP2 protein could be identified in some euryarchaeal groups (Törö et al. 2002). The TACK superphylum and the ASGARD group usually contain two additional Lsm proteins in addition to SmAP1, referred to as SmAP2 and 3 (Mura et al. 2013). *Saccharolobus solfataricus* (Sso) SmAP1 and 2 bind numerous RNA substrates including mRNAs, transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) and interact with various protein complexes as e.g. the exosome, which stimulates A-rich tailing of transcripts (Märtens et al. 2015, 2017a,b). Furthermore, Sso SmAP2 binds 3′UTRs of mRNAs in a sequence-specific manner and affects their stability (Märtens et al. 2017b). The role of SmAP3 is still unclear, but it is characterized by an extended C-terminal domain (Mura et al. 2003b).

Lsm-variants that show high conservation to the bacterial Hfq variant are only found in the family of the *Methanocaldococcaceae* and *Methanosaetaceae*. The Hfq-like SmAP from *Methanocaldococcus jannaschii* (MjHfq) forms homohexameric rings and shows comparable biochemical and biological properties to its *E. coli* Hfq counterpart (Nielsen et al. 2007). It can even functionally replace Hfq in *E. coli*.

Adaptation to changing environmental conditions is critical for the survival of prokaryotic cells and is also regulated at the posttranscriptional level (van Assche et al. 2015). However, the mechanisms of post-transcriptional regulation are poorly understood in Archaea. Given the structural similarity and RNA-binding capacity of SmAPs proteins, it is likely that they fulfil similar functions as bacterial Hfq proteins. A recent study in the halophilic archaeon *Halobacterium salinarium* indicated that SmAP1 together with antisense RNAs and RNases targets about 5.7 % of protein-coding genes (Kusebauch et al. 2023). However, at which step in the gene expression pathway SmAP1 binds RNAs and how regulation via this RNA-binding protein is executed, remained obscure.

We addressed these questions employing the SmAP protein from the archaeal organism *Pyrococcus furiosus* (Pfu) for which a genetic system is established (Grünberger et al. 2020; Waege et al. 2010). *P. furiosus* encodes a SmAP more akin to eukaryotic Lsm proteins (Reichelt et al. 2018). We created a mutant *P. furiosus* strain that constitutively expresses PfuSmAP1 genetically fused to a Strep-tag (Figure 1). This allowed us to co-purify nucleic acid and protein interaction partners that were further analysed by RNA-sequencing (RNA-seq) and mass spectrometry (MS) (Figure 1A). We demonstrate that PfuSmAP associates with various RNAs, among them messenger and antisense sRNAs (asRNAs). Notably, a large fraction of the associated RNAs were enriched for uridine. MS analysis identified various RNA polymerase (RNAP) subunits but also ribosomal proteins, the archaeal exosome and interestingly CRISPR-Cas (CRISPR: clustered regularly interspaced short palindromic repeats. Cas: CRISPR-associated proteins) proteins to be directly or indirectly associated with PfuSmAP1. Notably, we reveal that SmAP1 co-transcriptionally associates with RNAs. However, PfuSmAP1 does not directly interact with the RNA polymerase and, in contrast to the bacterial regulatory action of Hfq, does not influence transcriptional activity of the RNA polymerase *in vitro*.

2 Results

2.1 Purification of Strep-tagged SmAP1 and its co-purifying protein and nucleic acid interaction partners from *P. furiosus*

To decipher the role of SmAP1 in *P. furiosus*, we determined the protein and RNA interactome of SmAP1. To this end, we

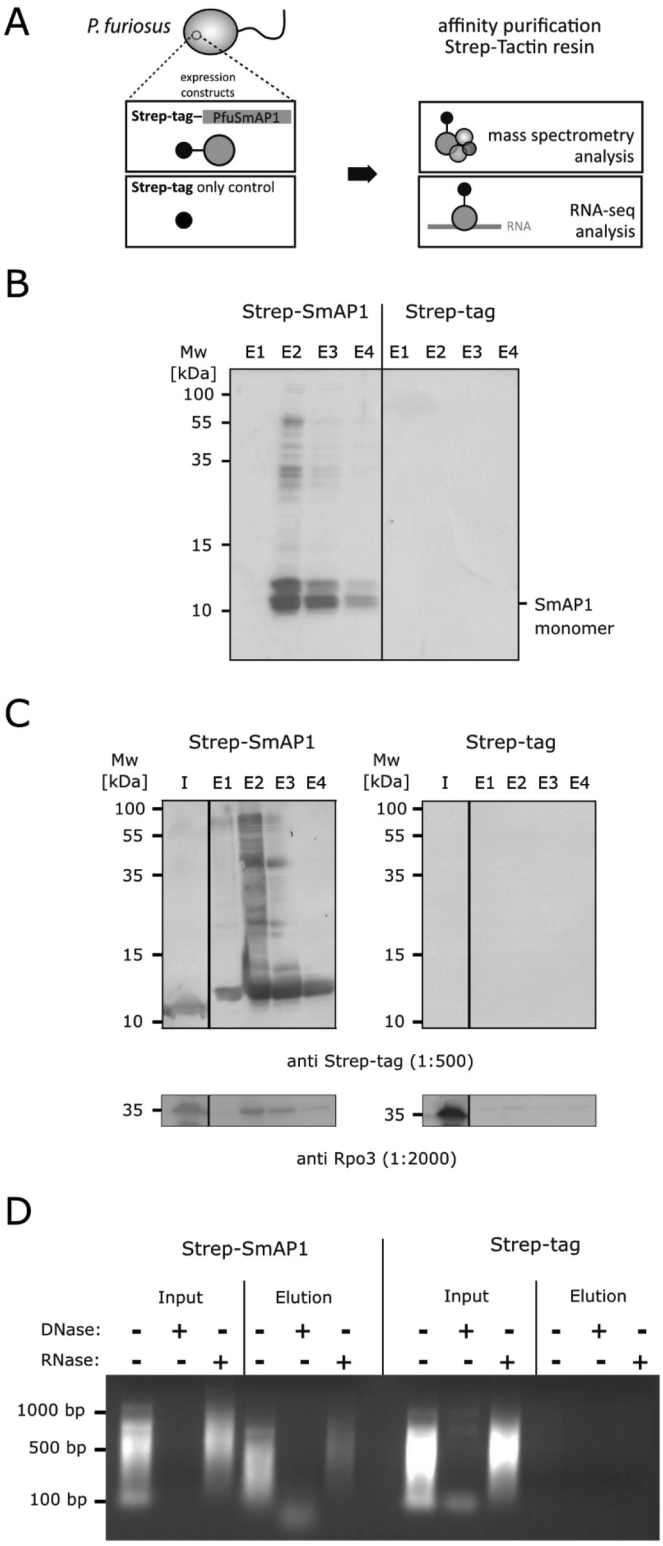

Figure 1: Purification of PfuSmAP1 from its cellular context elucidates RNA, DNA and protein interaction partners. (A) Experimental design illustrating the workflow to identify cellular interaction partners of the SmAP1 protein from *Pyrococcus furiosus* (PfuSmAP1). (B) SDS-PAGE analysis of affinity-purified elution (E) fractions using Strep-tactin resins loaded with crude extracts from a *P. furiosus* mutant constitutively overexpressing C-terminally Strep-tagged SmAP1 (Strep-SmAP1) or the Strep-tag peptide only (Strep-tag). (C) Western blot analysis showing the input (I) and elution (E) fractions after affinity purification. For detection, antibodies directed against the Strep-tag peptide (upper panel) and the RNA polymerase subunit Rpo3 (lower panel) were used. (D) Agarose-gel electrophoresis showing input and elution fractions. Input and elution fractions were subject to phenol-chloroform extraction and additionally treated with DNase or RNase before loading on the agarose gel.

took advantage of an established genetic system for *P. furiosus* that allows the overexpression of tagged fusion proteins from a shuttle vector (Grünberger et al. 2020; Waege et al. 2010). Accordingly, a mutant *P. furiosus* strain constitutively expressing C-terminally Strep-tagged SmAP1 was created for this purpose (Figure 1A). This allowed us to purify PfuSmAP1 and its protein and nucleic acid interaction partners directly from whole cell extracts via the Strep-tag.

As a control, a mutant *P. furiosus* strain constitutively expressing the Strep-tag peptide only was created. SDS-PAGE analysis of the purified sample revealed a prominent protein band that probably represents Strep-tagged PfuSmAP1 with a theoretical molecular weight of 9.6 kDa (Figure 1B). In addition, other bands corresponding to proteins with varying molecular weights are visible. These bands might represent partially denatured PfuSmAP1 multimers as well as interaction partners. Western blot analysis using an antibody directed against the Strep-tag peptide verified the successful purification of Strep-tagged PfuSmAP1 (Figure 1C).

In a next step we purified nucleic acids that associate with Strep-SmAP1 in the cellular context of *P. furiosus* and found both RNA and DNA substrates directly or indirectly bound by SmAP1 (Figure 1D). Control experiments with Strep-tag peptide only did not result in specific enrichment of proteins or nucleic acids (Figure 1B–D).

Moreover, to reveal its multimerisation state, we expressed PfuSmAP 1 recombinantly in a Δhfq *E. coli* strain (Märtens et al. 2015) to avoid co-purification of Strep-SmAP1 and *E. coli* Hfq as heterocomplexes (Nielsen et al. 2007). During the purification, the samples were rigorously treated with DNase and RNase to obtain highly pure SmAP1 elution fractions. Subsequent SDS PAGE analysis revealed successful purification of monomeric Strep-PfuSmAP1 (Supplementary Figure 1A). To determine if this monomeric Strep-PfuSmAP1 forms multimeric assemblies, we performed mass photometry measurements (Supplementary Figure 1B). Indeed, in absence of nucleic acids, PfuSmAP1 assembles into heptameric complexes (76 %; 72 ± 10 kDa). Moreover, dimers of heptamers (14 %; 147 ± 12.3 kDa) were detected as a minor fraction. The formation of heptamer dimers agrees with structural studies of SmAP1 from *Pyrococcus abyssi* (Thore et al. 2003).

2.2 SmAP1 interacts with various RNAs containing an U-rich motif *in vivo*

Using deep sequencing, we analysed the identity of co-eluted RNAs. We identified 199 different RNA binding partners and mapped them based on the improved re-annotation of the *P. furiosus* genome (Figure 2A; Table S1) (Grünberger et al. 2019). SmAP1 mainly binds mRNAs (n = 130). The binding site is often located at the 3′ end of the mRNAs and we detected an overlap for 39 regions with an ANNOgesic-annotated 3′ UTR (Figure 2B; Table S1).

In addition, SmAP1 binds many RNAs (n = 50), which are oriented in antisense to an mRNA (asRNAs). For eleven of these putative asRNAs, the presence of a corresponding antisense transcript was detected in the comprehensive transcriptome study of Grünberger et al. (2019). SmAP1 also binds five ncRNAs located in non-coding regions and seven ncRNAs located within a mRNA. Finally, associations of SmAP1 with four CRISPR RNAs (crRNAs), two overlap small ORFs (sORFs) and only one tRNA could be detected. Motif analysis revealed a gapped U-rich motif in 97 % of all binding regions (Figure 2C; Table S1). Deep sequencing analysis of co-purified DNA fragments did not reveal any enrichment of specific loci (data not shown).

The binding of PfuSmAP1 to numerous mRNAs and asRNAs suggests an involvement in the regulation of various cellular processes and pathways. To analyse this in more detail, the corresponding mRNAs were classified according to arCOG categories (Makarova et al. 2015) (Figure 2D). This revealed that PfuSmAP binds and potentially affects mRNAs encoding proteins that fulfil tasks in e.g. amino acid transport and metabolism (E), carbohydrate transport and metabolism (G), translation, ribosome structure and biogenesis (J), transcription (K), cell wall/membrane/envelope biogenesis (M) and others. A significant enrichment (p < 0.05) was detected for mRNAs of the arCOG category amino acid transport and metabolism (E). For mRNAs that presumably bind PfuSmAP1 via an asRNA, the arCOG category cell wall/membrane/envelope biogenesis (M) was significantly enriched (Table S1).

2.3 SmAP1 interacts *in vivo* with various proteins associated to specific biological processes and multiprotein complexes (RNPs)

Previous studies showed the interaction of SmAP proteins with numerous proteins similarly to their eukaryotic (Sm and Lsm) and bacterial (Hfq) counterparts (Butland et al. 2005; Ho et al. 2002; Krogan et al. 2004; Uetz et al. 2000). To shed light on the SmAP1 protein interactome in *P. furiosus*, we identified proteins that co-purify with Strep-tagged SmAP1 from cell lysate by mass spectrometry (Figure 3). In addition to untreated samples, a second set of samples was rigorously treated with DNase and an RNase-mix before purification. Purifications using crude extracts derived from the Strep-tag only mutant served as control to identify non-specifically extracted proteins. 95 specifically co-purifying proteins were identified in the untreated samples (Figure 3A, Table S2). Nuclease treatment decreased the number of proteins to 23 (Figure 3B, Table S2). Interactions from untreated samples included proteins involved in the RNA metabolism of the cell

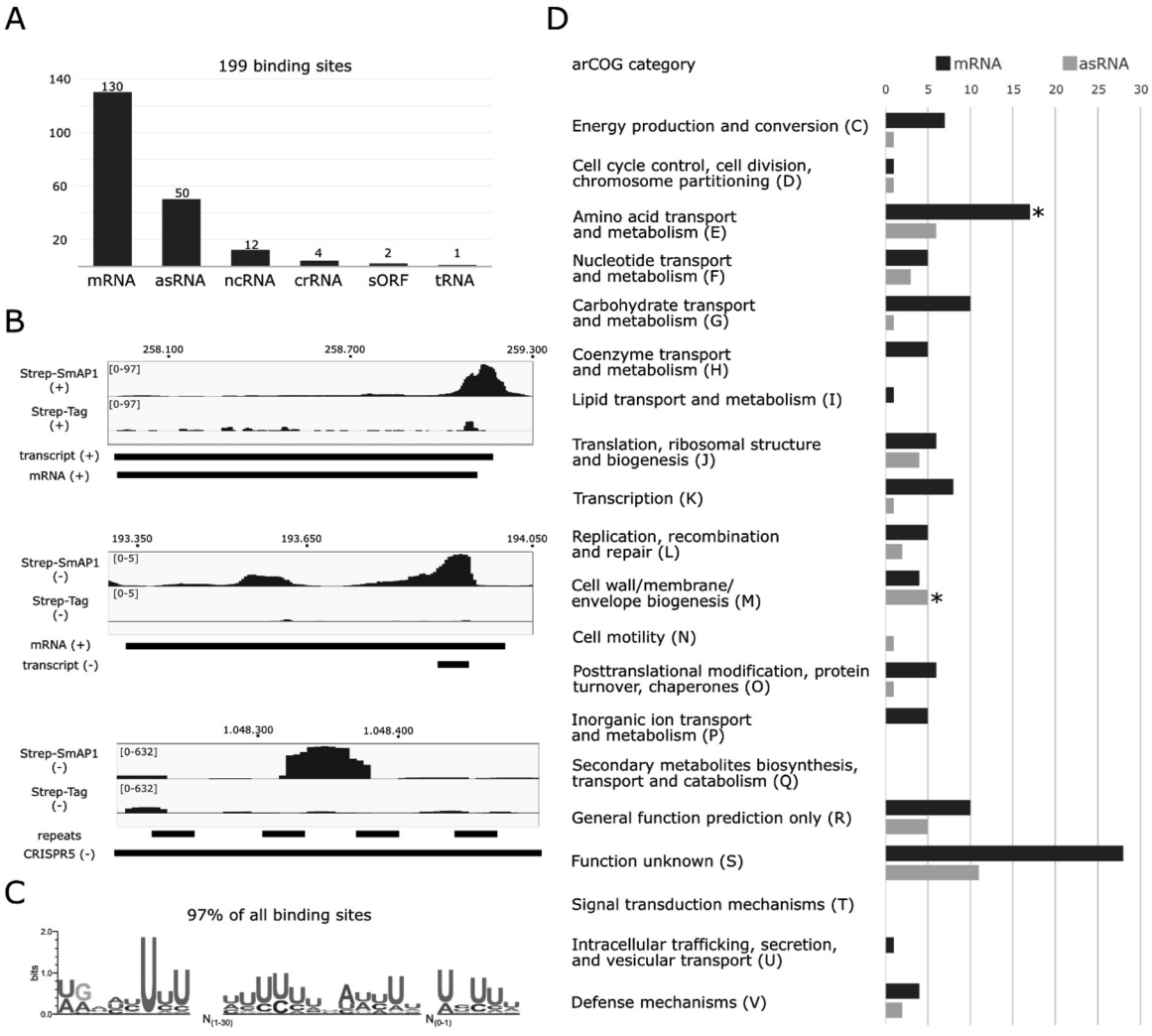

Figure 2: SmAP1 interacts with RNAs containing a U-rich motif *in vivo*. (A) Bar chart showing the variety of RNA types associated with PfuSmAP1 *in vivo*. (B) Strand-specific visualisation of mapped RNA-seq reads of PfuSmAP1 (Strep-SmAP1) or the Strep-tag peptide only (Strep-tag) at specific chromosomal loci using the Integrated Genomics Viewer (Robinson et al. 2011). Annogesic derived annotations (mRNA, transcript, crRNA and CRISPR repeats) are shown as black bars. (C) WebLogo three representation (Crooks et al. 2004) of the gapped U-rich PfuSmAP1 binding motif indentified using Glam2. (D) Enrichment analysis of archaeal clusters of orthologous genes (arCOGs) of mRNAs directly associated with PfuSmAP1 (dark blue bar) and indirectly associated via an asRNA, which interacts with PfuSmAP1 *in vivo* (light blue bar). arCOG groups with a p-value < 0.05 are considered significantly overrepresented and are highlighted by a black star.

including the exosome, ribosome, RNA polymerase and components of two CRISPR Cas systems (type I-A and III-B) of *P. furiosus*. In addition, we identified proteins involved in processing or modification of tRNAs (e.g. FlpA), rRNAs (e.g. FAU-1) and mRNAs (e.g. transcription termination factor aCPSF1). Other protein interactions were involved in DNA replication and DNA repair. An analysis based on arCOG categories (Makarova et al. 2015) revealed significant enrichments of the categories translation, ribosomal structure and biogenesis (J); transcription (K); replication, recombination and repair (L) and defence mechanisms (V) (Table S2). Nuclease treatment led to the loss of some interaction partners including RNA polymerase subunits. In addition, several co-purifying proteins, which are involved in RNA processing and modification, were no longer detected suggesting that these interactions are mediated by RNAs.

2.4 Association of SmAP1 with the RNA polymerase complex is RNA-dependent

In Bacteria, under certain conditions Hfq acts as an anti-termination factor binding mRNA co-transcriptionally and preventing pre-mature binding of Rho (Sedlyarova et al. 2016). To gain more insights into a putative interaction

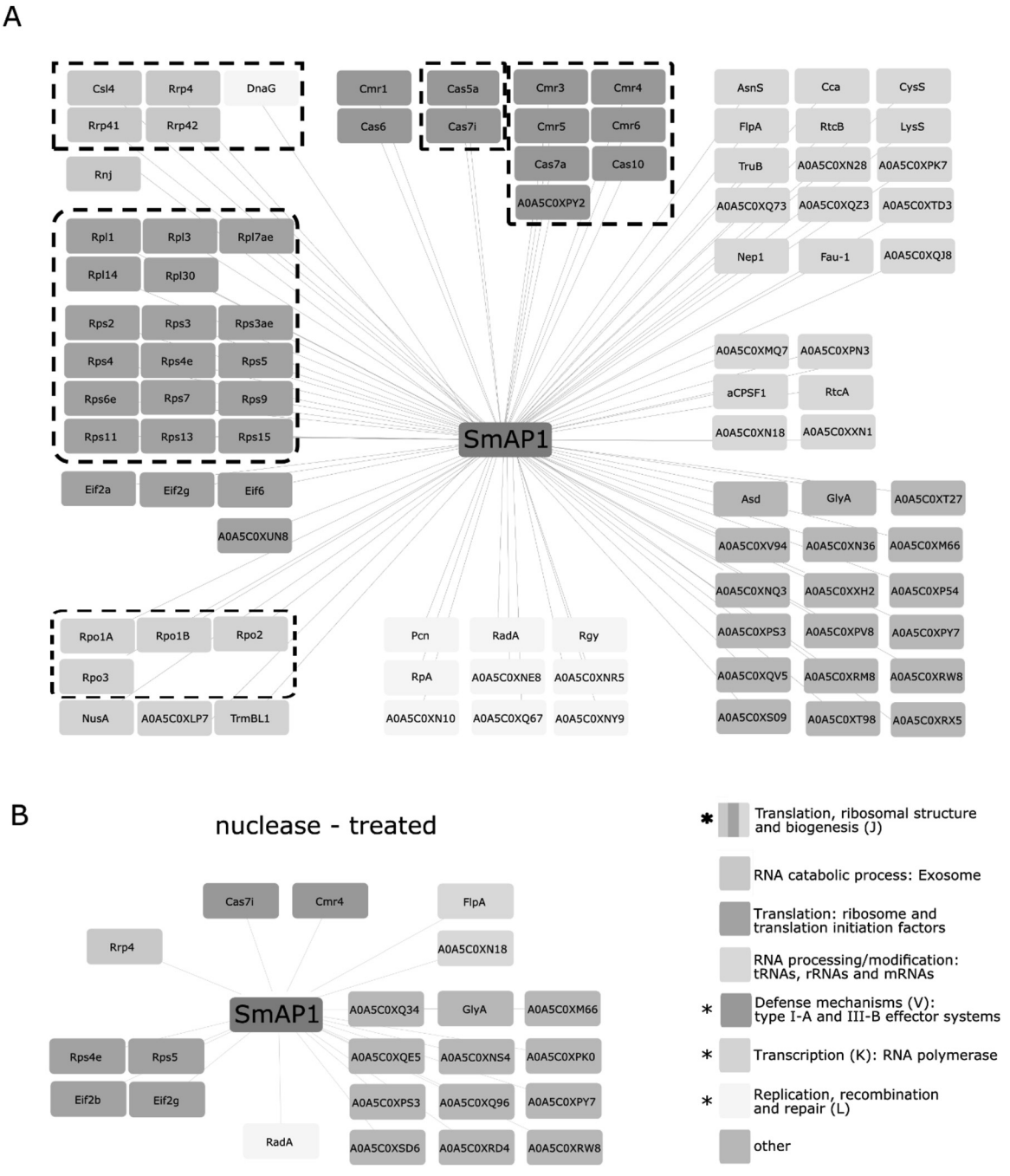

Figure 3: SmAP1 interacts with various proteins in an RNA-dependent and RNA-independent manner *in vivo*. Representation of PfuSmAP1 interaction networks using Cytoscape (Saito et al. 2012) *in vivo*. Interaction partners were identified by mass spectrometry analysis after purification using untreated (A) or DNase and RNase treated (B) crude extracts from a *P. furiosus* mutant constitutively overexpressing C-terminally Strep-tagged PfuSmAP1. Crude extracts from a *P. furiosus* mutant constitutively overexpressing the Strep-tag peptide only served as background control. Identified proteins were classified using the UniProt Knowledgebase (UniProt Consortium 2023) and corresponding genes using the archaeal clusters of orthologous genes (arCOGs) (Makarova et al. 2015). Co-purified proteins that form a functional multiprotein complex like the exosome, ribosome, RNA polymerase and the CRISPR Cas effector complexes of the Type-I-A and III-B are highlighted by dotted rectangles. DnaG was placed within the exosome complex according to (Hou et al. 2014; Witharana et al. 2012). Rnj was placed associated to the exosome according to (Phung et al. 2020). Significantly ($p < 0.05$) overrepresented arCOG categories (star regular font: untreated only; star bold font: untreated and treated) are highlighted by colours as shown in the Figure.

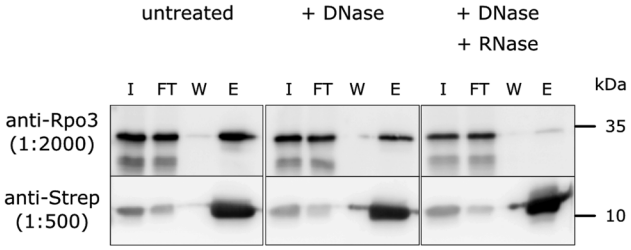

Figure 4: Association of SmAP1 with the RNA polymerase complex is RNA-dependent. Western blot analysis of samples derived from purifications of SmAP1 (Input, I; Flowthrough, FT; Washing, W; Elution, E) from crude extracts from a *P. furiosus* mutant constitutively overexpressing C-terminally Strep-tagged SmAP1. Crude extracts were untreated, only treated with DNase or treated with DNase and RNase, respectively, before purification of the PfuSmAP protein. For detection, an antibody raised against the RNA polymerase subunit Rpo3 (upper panel) and an antibody specific for the Strep-tag peptide (lower panel) were used.

between PfuSmAP1 and the archaeal RNAP, we performed additional co-purification experiments. These experiments aimed to test whether SmAP1 directly interacts with the RNAP or indirectly via RNA that emerges from the RNAP. To this end, we purified Strep-SmAP1 from (i) untreated, (ii) DNase treated or (iii) DNase and RNase treated cell lysates. Subsequently, we performed SDS PAGE analysis followed by Western blotting with antibodies directed against the RNAP subunit Rpo3 and the Strep-tag peptide to verify co-purification of RNAP with Strep-tagged PfuSmAP1 (Figure 4). All conditions led to the purification of comparable amounts of Strep-tagged PfuSmAP1. DNase treatment only had a minor effect on co-purification of the RNAP, whereas additional RNase treatment resulted in a significant decrease in the detectable Rpo3 signal.

2.5 *P. furiosus* SmAP1 binds U-rich RNAs *in vitro* and positively effects transcription

Co-transcriptional binding of SmAP1 to emerging transcripts suggests a role in transcriptional regulation. Hence, in a next step, we analysed the effects of PfuSmAP1 association with RNAs during transcription elongation in more detail. Therefore, we chose a transcript that contains a putative U-rich SmAP binding site in the gene PFDSM3638_05070, which encodes a putative phosphate transport regulator (PhoU) (Figure 5A). To verify binding of PfuSmAP1 to RNAs with this U-rich motif, we performed EMSAs with a 38 nt HEX-labelled version of this RNA stretch (Figure 5B). PfuSmAP1 readily formed a complex with this RNA that migrated slower than the free RNA in a native gel. We observed a shift of more than 50 % of the RNA into the PfuSmAP1-RNA complex at a SmAP1 concentration of 25 nM (ratio PfuSmAP1 heptamer:RNA = 2.5:1). This reveals a high affinity of PfuSmAP1 for this RNA in the nanomolar range. A control version of this RNA substrate with nine cytosine bases instead of the uridine bases (Figure 5A, underlined bases) did result in a shift of more than 50 % of free RNA at a concentration of 200 nM PfuSmAP1 (ratio PfuSmAP1 heptamer:RNA = 20:1).

Next, we monitored the effect of SmAP1 on transcription. We chose promoter-directed transcription assays using the strong promoter of the histone A1 gene (*hpyA1*) to drive transcription of the partical *phoU* gene (Figure 5A and C). It is noteworthy that the U-rich sequence potentially also acts as an intrinsic termination sequence. Earlier studies demonstrated that poly-dT stretches of five or more nucleotides in the template DNA induce termination of transcription *in vitro* (Hirtreiter et al. 2010; Santangelo and Reeve 2006; Spitalny and Thomm 2008). Analysis of synthesized transcripts showed that the oligo-dT6 stretch only partially induces termination resulting in a 135 nt transcript and that most of the transcripts correspond to full-length transcripts (395 nt). This indicates efficient readthrough of the putative terminator. The overall pattern of transcripts did not change when the transcription reaction was carried out in the presence of SmAP1. However, the overall transcription yield increased slightly when SmAP1 was added to the transcription reaction (125 % ± 5 % for the run-off at a SmAP1 concentration of 400 nM) (Figure 5C).

3 Discussion

Sm-like proteins are widely distributed in Archaea, but their functional roles are still poorly understood. Recombinant expression and *in vitro* characterization of archaeal SmAPs have been performed previously, providing insights into the structural organization of archaeal SmAPs and their RNA binding capacities. In agreement with these studies, we found that recombinant PfuSmAP1 forms homoheptamers *in vitro* (Achsel et al. 2001; Fischer et al. 2010; Kilic et al. 2005; Thore et al. 2003; Törö et al. 2002), whereas bacterial Hfq proteins and the Hfq-like protein from *M. jannaschii* assemble into homohexamers (Nielsen et al. 2007; Schumacher et al. 2002). A major challenge in understanding the physiological binding partners and functional roles of SmAPs is the lack of genetic systems for a wide range of archaeal species. So far, this limits *in vivo* studies to the halophilic and mesophilic Archaeons *H. volcanii* and *Halobacterium salinarum* as well as the hyperthermophilic Archaeon *S. solfataricus* (Fischer et al. 2010; Kusebauch et al.

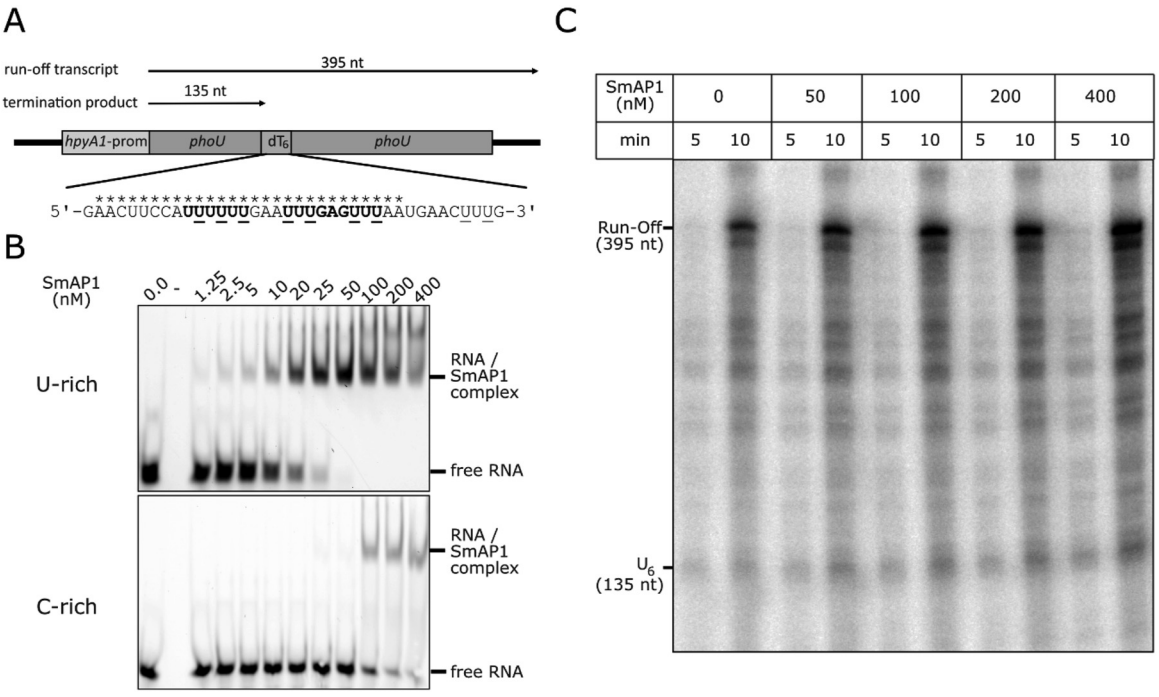

Figure 5: PfuSmAP1 binds U-rich RNAs *in vitro* and positively effects transcription. (A) Illustration of DNA template used for promoter-dependent *in vitro* transcription assays. The sequence of the *hpyA1* promoter and the first transcribed 25 nt (light blue) (Spitalny and Thomm 2008) was fused to parts of the *phoU* gene (PFDSM3638_05070; dark blue) containing a PfuSmAP1 binding region and a dT-rich region (orange). The corresponding U-rich RNA sequence stretch is shown in detail and was also used in EMSA experiments. Length of the expected intrinsic termination and run-off transcription product are shown. The SmAP1 RNA binding motif is highlighted by stars. (B) EMSAs using HEX-labelled RNAs and increasing amounts of recombinant PfuSmAP1. As RNAs, either the U-rich RNA shown in (A) was used (upper panel) or an RNA in which nine U nucleotides were replaced (underlined in A) by Cs (C-rich, lower panel). Complexes were separated using native gel electrophoresis. (C) Promoter-dependent transcription assays using the DNA template shown in (A). Increasing amounts of recombinant PfuSmAP1 were used in the transcription assay. Samples were taken after a reaction time of 5 and 10 min at 80 °C. RNA transcripts were extracted by phenol-chloroform extraction and separated on a denaturating 15 % PAA gel. Intrinsic termination (U_6) and run-off signals are highlighted.

2023; Maier et al. 2015; Märtens et al. 2015, 2017a). In this study, we took advantage of an established genetic system for the hyperthermophilic Euryarchaeote *P. furiosus*. We applied the Strep-tag system to purify SmAP1 from its cellular context to decipher both its RNA and protein interactome and expand physiological studies to hyperthermophilic Euryarchaea.

We found that PfuSmAP1 mainly binds to mRNAs and a significant number of asRNAs and ncRNAs. Given that we detected 130 mRNA binding partners and that 1982 protein-coding genes are found in the genome of *P. furiosus* (Grünberger et al. 2019), PfuSmAP1 binds 6.6 % of all protein coding transcripts. This is in agreement with studies of SmAPs in *S. solfataricus* and *H. salinarum*, which were shown to preferentially interact with mRNAs and asRNAs (Kusebauch et al. 2023; Märtens et al. 2015). For example in *H. salinarum,* SmAP1 binds to 15 % of the protein coding transcripts (Kusebauch et al. 2023). Our data revealed a bias for PfuSmAP1 binding sites in the 3′-UTR of mRNAs. This suggests functional similarity to SsoSmAP2, which binds a specific motif in mRNA 3′ UTRs (Märtens et al. 2017b). SsoSmAP2 induces polyadenylation of mRNAs and thereby affects their stability (Märtens et al. 2017b). Other SmAPs like e.g. HvoSmAP mainly bind small, putatively regulatory RNAs (sRNAs) (Fischer et al. 2010).

We furthermore explored the sequence specificity of PfuSmAP1. Many studies demonstrated that other SmAPs efficiently bind poly(U) RNAs *in vitro*. However, SsoSmAP2 favours GC rich sequences. In case of PfuSmAP1, we found a preference for U-rich sequences and detected a gapped poly(U)-binding motif in mRNAs, asRNAs and ncRNAs that interact with PfuSmAP1.

Exploring the PfuSmAP1 RNA interactome, we did not detect interactions with tRNAs, snoRNAs or H/ACA box and C/D box RNAs, which is in contrast to earlier studies (Fischer et al. 2010; Märtens et al. 2017a; Törö et al. 2001). However, the protein interactome that we found in our study suggests that SmAP1 plays a role in binding these RNA classes. This is corroborated by the fact, that most of the corresponding protein interactions were lost after DNase and RNase

treatment. UV irradiation used in other studies might have stabilized otherwise transient interactions, which might explain the difference to our data (Fischer et al. 2010). Notably, we found many asRNAs to interact with PfuSmAP1. This suggests an RNA chaperone activity of PfuSmAP1 that enables interactions between non-coding RNAs and mRNAs. This chaperone activity by Hfq is involved in post-transcriptional regulation of gene expression by restricting access to the ribosomal binding site, or to RNases thereby regulating RNA stability (Møller et al. 2002; Sledjeski et al. 2001; Wassarman et al. 2001; Zhang et al. 2002, 2003). In addition, Hfq also plays a role in the pairing of cis-encoded asRNAs as shown for example for the *E. coli* Tn*10*/IS*10* antisense system (Ross et al. 2013). A potential RNA chaperone activity is supported by the location of the PfuSmAP1 binding sites in mRNA 3′-UTRs. These regions often serve as regulatory hotspots in bacterial and archaeal posttranscriptional regulation via small RNAs (Dar et al. 2016; Menendez-Gil and Toledo-Arana 2020). SmAP1 RIP-seq data from studies with *H. salinarum* in combination with RNAseq data of wildtype and an RNase knockout strain as well as ribosome profiling and proteomics data delivered further support for a role of SmAP1 in cooperation with asRNAs and RNases in post-transcriptional regulation (Kusebauch et al. 2023). This data revealed that at least 7 % of all protein-coding genes in *H. salinarum* are subject to posttranscriptional regulation exerted by SmAP1, asRNAs and RNases. Based on this data, it was suggested that SmAP1 and asRNAs might recruit RNase_2099C to transcripts resulting in target cleavage (Kusebauch et al. 2023) but this hypothesis has not yet been biochemically validated. The mode of a putative interaction between perfectly base-pairing asRNAs and mRNAs mediated by PfuSmAP1 remains to be elucidated. Furthermore, it is unclear how RNases are recruited to this ribonucleoprotein particle. Our analysis did not allow the detection of two RNAs that are simultaneously bound by SmAP1, which are likely to be a regulatory RNA pair.

An additional noteworthy aspect is the enrichment of crRNAs, and more specifically U-rich crRNAs. This has also been observed for SsoSmAP1 and SsoSmAP2 and for bacterial Hfq proteins (Boudry et al. 2021; Märtens et al. 2015; Trouillon et al. 2022). However, the functional implications are not understood yet.

In summary, PfuSmAP1, like many other Sm and Sm-like proteins, binds a variety of cellular RNAs recognizing U-rich sequences. Remarkable and unique for PfuSmAP1 are (i) the preferential binding of asRNAs and (ii) the complex gapped binding motif. The latter may indicate that (i) PfuSmAP1 can bind two independent RNAs that both possess a U-rich stretch, (ii) that RNAs mediate the formation of SmAP multimers *in vivo* as demonstrated for *Archaeoglobus fulgidus*, *M. thermophilus* and *Pyrococcus aerophilum in vitro* (Mura et al. 2003a), or (iii) that binding of multiple SmAP1 molecules on the same RNA with multiple poly(U) stretches is possible. Indeed, arrays of poly(U) sequences that promote transcription termination are found at the 3′ end of archaeal RNAs (Dar et al. 2016; Yue et al. 2020).

Also protein interaction partners identified for PfuSmAP1 are conserved for other SmAPs (Fischer et al. 2010; Märtens et al. 2017a). These include the exosome and DnaG, which are critically involved in 3′–5′ RNA degradation and/or polyadenylation in most Archaea except for halophiles (Phung et al. 2020; Portnoy and Schuster 2006). Another common theme is the interaction of SmAPs with rRNA- and tRNA-modifying enzymes, ribosomal proteins and translation initiation and elongation factors. Most of these interactions appear to be mediated by RNAs rather than being direct interactions. Fibrillarin-like rRNA/tRNA 2′-O-methyltransfere (FlpA) as an exception remains associated with PfuSmAP after RNase digestion of the samples. FlpA is a homologue of Nop1 (yeast) and the human fibrillarin, respectively, that are components of the snoRNA-guided methylation complex, which is also conserved in Archaea (Breuer et al. 2021). SnoRNAs and Nop56, another component of this methylation complex, were identified as interaction partners of SmAP1 and SmAP2. This strengthens the assumption, that SmAPs are involved in rRNA modification processes and thereby influence ribosome biogenesis.

Direct interactions of SmAPs with CRISPR Cas systems and crRNAs were so far solely detected in *S. solfataricus.* In *P. furiosus*, SmAP1 interacts with two of three CRISPR-Cas-effector complexes (I-A and III-B, according to Makarova et al. [2020]). Whether this interaction is truly a direct interaction or mediated by a tightly bound and thereby protected crRNA remains unknown. We found crRNAs associated with PfuSmAP1. This was also observed in case of SsoSmAP1, SsoSmAP2 and bacterial Hfq proteins (Boudry et al. 2021; Trouillon et al. 2022). However, the functional implications are not understood yet.

An association of SmAPs with RNAP has only been shown for *H. volcanii* so far (Fischer et al. 2010). In *P. furiosus*, this interaction depends on the presence of RNA suggesting co-transcriptional binding of SmAP1 during transcription elongation. During this step the nascent RNA is sufficiently long to allow an association of PfuSmAP1. Furthermore, the elongation provides the time window for the simultaneous binding of RNAP and PfuSmAP to the same RNA before the RNAP dissociates from the transcript during termination. This assumption is further supported by the fact that we were also able to detect an interaction of PfuSmAP1 with the transcription elongation factor NusA. In

Bacteria, Hfq-RNA complexes interact with nascent transcripts close to elongation complex allowing a faster recognition of target RNAs (Rodgers et al. 2023) thereby exerting its regulatory effects at the earliest possible opportunity (Kambara et al. 2018). In addition, *E. coli* Hfq positively regulates the expression of sigma factor RpoS by blocking access of the termination factor Rho to the long 5′ UTR of the RpoS transcript (Sedlyarova et al. 2016). This way, Hfq acts as *bona fide* antitermination factor. Rho is not conserved in Archaea, but we found the archaeal termination factor aCPSF1 as interaction partner of PfuSmAP1. aCPSF1 terminates transcription by cleaving the transcript at the 3′ end (Sanders et al. 2020). Interestingly, aCPSF1 requires U-rich regions as binding sites (Li et al. 2021; Sanders et al. 2020). In our study, the presence of SmAP1 in cell-free transcription assays did not affect the overall transcriptional pattern, nor did it prevent RNAP pausing. However, as found for *E. coli* Hfq, an overall increase in transcriptional output was observed (Sukhodolets and Garges 2003). Since both RNA-binding proteins do not directly interact with the RNAP, the stimulatory effect on transcript elongation could be due to a stabilizing effect of the RNA-binding protein on newly synthesized RNAs. The fact that both, aCPSF1 and SmAP1, bind to U-rich sequences potentially opens an alternative pathway for transcriptional and posttranscriptional regulation (Weixlbaumer et al. 2021) that does not involve the recruitment of RNases. It would be feasible, that SmAP binds the 5′ UTR of mRNAs. The interaction of SmAP with ribosomal proteins could promote recruitment of ribosomes to the mRNA supporting the coupling of transcription and translation in Archaea. Ribosome binding could prevent aCPSF1 from interacting with the nascent RNA, thereby preventing premature termination. Antitermination by PfuSmAP1 could also be a result of a direct competition between SmAP1 and aCPSF1 for poly(U) stretches at the 3′ end of the RNA.

Taken together, our study confirms the central role of SmAPs in archaeal RNA metabolism, but also reveals previously unrecognized interactions with defence systems and the transcriptional apparatus. On the one hand, these novel findings need to be analysed in more detail to gain further insights into transcriptional and posttranscriptional regulation mediated by SmAP, but it would also be worthwhile to investigate the functions of Sm-like proteins in more diverse groups within the archaeal domain of life. The repertoire of Sm-like proteins ranges from at least one (SmAP1 or Hfq-like SmAP) in the DPANN group and Euryarchaeota to up to three (SmAP1, 2 and 3) and even more in the TACK superphylum and Asgard group (Mura et al. 2013; Reichelt et al. 2018). This represents a hitherto untapped reservoir of diversity, which certainly still harbours many previously unknown functions of archaeal Sm-like and Hfq-like proteins.

4 Materials and methods

4.1 Strains, plasmids, and primers

All strains, plasmids and primers used in the study are listed in Table S3.

4.2 Construction of *P. furiosus* overexpression strains and growth conditions

For the construction of *P. furiosus* strains expressing Smap1 with the Strep-tag® II fused to the C-terminus or with the Strep-tag® II peptide only, the modified genetic system for *P. furiosus* DSM3638, based on selection via agmatine auxotrophy as described in Grünberger et al. (2020) was used.

Based on plasmid pMUR310 (Grünberger et al. 2020), an overexpression shuttle vector system was constructed that allows constitutive expression of a protein of interest via the *gdh* promoter (PCR amplified from *Thermococcus kodakarensis* genomic DNA using primers RPA_pYS13_GA_F and pYS13_GA-R) and the hpyA1 terminator (PCR amplified from *P. furiosus* genomic DNA using primers PYS14_GA_F and pYS13_RPA_GA_R) regions. Both regions were fused upstream and downstream of the gene of interest as described in Waege et al. (2010) and cloned into pMUR310 via the EcoRV restriction site. One of these overexpression plasmids was used for PCR amplification of the plasmid backbone using primers pYS14Expgdh_F and pYS14_GA_F. The *smap1* gene was PCR amplified from *P. furiosus* genomic DNA using primers GAp14PF1542FW and GAp14PF1542RW. Both PCR products were assembled using Gibson Assembly (Gibson et al. 2009), resulting in pMUR433. To insert the Strep-tag® II sequence, a Phusion® High-Fidelity DNA Polymerase mutagenesis PCR (New England Biolabs) was performed using the primer pYS14_GA_F and the phosphorylated primer PF1542cstrep2RWP to amplify the whole plasmid. The PCR product was ligated using T4 ligase enzyme (New England Biolabs) and transformed into *E. coli* DH5α cells (New England Biolabs), resulting in pMUR435. The final construct including the correct insertion of the Strep-tag® II sequence was verified by sequencing (Microsynth). Finally, to construct an overexpression shuttle vector containing only the Strep-tag® II sequence, Phusion® High-Fidelity DNA Polymerase mutagenesis PCR (New England Biolabs) was performed using the pYS14_GA_F primer and the phosphorylated p14strepfw primer. The PCR product was ligated with T4 ligase enzyme and transformed into *E. coli* DH5α cells (New England Biolabs), resulting in pMUR544. The correct removal of the *smap1* sequence was verified by sequencing (Microssynth). 1 µg of the circular plasmids pMUR435 or pMUR544 was transformed into MURPf37 as described (Grünberger et al. 2020; Kreuzer et al. 2013; Waege et al. 2010). Selection was carried out in 1/2 SME liquid medium without agmatine sulphate and inosine+guanosine at 85 °C for 12 h. Pure cultures of the mutants MURPf80 (pMUR435) and MURPf82 (pMUR544) were obtained by plating the cells on solidified medium. Plasmid stability was verified by re-transformation into *E. coli* DH5α and DNA sequencing of purified plasmids (Microsynth). Final *P. furiosus* mutants could be grown without agmatine sulphate and I+G supplementation.

MURPf80 (PfuSmAP1-Strep-tag) and MURPf82 (Strep-tag only) were cultivated in a 100 L biofermentor under anaerobic conditions at 95 °C in nutrient-rich medium based on 1/2 SME medium (Fiala and Stetter 1986; Reichelt et al. 2016) supplemented with 0.1 % (w/v) starch, 0.1 % (w/v) yeast extract, and 0.1 % (w/v) peptone. Cells were harvested after reaching the late-exponential phase (cell density of 1.5×10^8 cells per ml) and stored at −80 °C.

4.3 Identification of RNAs that co-purify with via RNA-seq

0.2 g of *P. furiosus* cell pellet was resuspended in 0.8 ml Strep100 buffer (10 mM Tris/HCl pH 8.0, 100 mM NaCl, 10 % (v/v) glycerol). The cell suspensions were sonicated for 12 × 3 min using a Bandelin Electronic™ Sonopuls™ HD 2070 homogenizer (cycle: 60 %, power 60 %). After centrifugation for 60 min at maximum speed (4 °C), the supernatant was applied to a gravity column containing 500 µl of Strep-Tactin® Sepharose® resin (iba). After 10 wash steps (1 ml each) with Strep100 buffer, elution was performed with 1× Elution Buffer E (IBA Lifesciences). The following sequential elution volumes were used: 400 µl, 700 µl, 400 µl, 400 µl. The success of the purification was analysed by SDS-PAGE and Western blotting. Co-purified nucleic acids from elution step E2 were used for deep sequencing. The experiment was replicated three times using cell pellets from MURPf80 (PfuSmAP1-Strep-tag mutant) and MURPf82 (Strep-tag only mutant). For RNA sequencing, half of the eluate (350 µl of fraction E2) was treated with 2U TURBO™ DNase (Invitrogen™) and the other half for DNA sequencing after treatment of the sample with 100 µg/ml RNase A (ThermoFisher) for 60 min at 37 °C. After phenol-chloroform extraction followed by ammoniumacetate/ethanol precipitation, RNA and DNA samples were analysed by agarose gel electrophoresis. Sequencing was performed by the Core Unit SysMed at the University of Würzburg. RNA quality of the six RNA samples was checked using a 2100 Bioanalyzer with the RNA 6000 Pico kit (Agilent Technologies). cDNA libraries suitable for sequencing were prepared from 130 ng of fragmented RNA treated with T4 PNK for phosphorylation/dephosphorylation and RppH for decapping followed by NEBNext® Multiplex Small RNA Library Prep (New England Biolabs). 15 PCR cycles and 30 s elongation time were used for amplification of the library. The final libraries had an average size of 155 bp. DNA libraries from the six DNA samples were prepared from 50 ng of fragmented DNA. The NEBNext Ultra II DNA Lib kit was used for library preparation. For amplification a PCR reaction with four amplification cycles was run. The final libraries had an average size of 500–530 bp. Both types of libraries were quantified by QubitTM dsDNA HS Assay Kit 3.0 Fluometer (ThermoFisher) and quality was checked using a 2100 Bioanalyzer with High Sensitivity DNA kit (Agilent Technologies) before pooling. Sequencing of pooled libraries, spiked with a 5 % PhiX control library, was performed in single-end mode with 150 nt read length on the NextSeq 500 platform (Illumina) with a Mid Output Kit. Demultiplexed FASTQ files were generated with bcl2fastq2 v2.20.0.422 (Illumina).

Reads in FASTQ format were quality/length/adaptor trimmed using trimmomatic in single-end-mode (Bolger et al. 2014). We allowed for a minimum length of 15 bases and a cut-off Phred score of 20, calculated in a sliding window of four bases. As adapter input the NEB_R1 sequence (5'AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC3') was provided. We used Bowtie 2 (Langmead et al. 2009) to map the reads to the 2019 updated version of the *P. furiosus* genome (Grünberger et al. 2019). Read count and mapping statistics are shown in Table S1. Next the sorted BAM files of the PfuSmAP1-Strep-Tag samples were analysed by Homer (v3.12) (Heinz et al. 2010) to define read-enriched regions in a strand separated manner. A tag directory was created and used as input to run findPeaks (-style factor -o auto -gsize 2e6 -strand separate -size 150 fragLength 75 -localSize 10000). This gave in total 613 read-enriched intervals on the plus and minus strand which could be detected in all three replicates (Table S1). We used Salmon to quantify read abundances for all six RNA-seq samples (3 × PfuSmAP1-Strep-Tag and 3 × Strep-tag only) within the Homer derived intervals (Patro et al. 2017) followed by a differential gene expression analysis using the DESeq2 pipeline (Love et al. 2014). Homer-enriched regions were considered as PfuSmAP1 binding site if they showed a log2 FC > 0.0 and an adjusted *p*-value of <0.01 (Table S1).

The found intervals ($n = 199$) were mapped to the following features of the Annogesic pipeline: mRNA, pseudogene, pseudogenic exon, sORF, rRNA tRNA, crRNA, ncRNA, transcript, 5UTR, 3UTR (Table S1). SmAP1 binding sites overlapping more than one feature were curated manually. Moreover the intervals were analysed using GLAM2 (Gapped Local Alignment of Motifs) to detect preferred binding motifs within the sites (Frith et al. 2008) using the MEME Suite (Bailey et al. 2015). Additionally, functional enrichment analysis based on the Archaeal Clusters of Orthologous Genes (arCOG) classification was performed as described previously (Knüppel et al. 2021). Briefly, arCOGs for *P. furiosus* were retrieved from Makarova et al. (2015) and gene set enrichment analysis was performed with the goseq package in R, which accounts for gene lengths bias (Young et al. 2010). For comparison all genes of the *P. furiosus* genome annotation were used. Next, *p*-values for over-representation of arCOGs were calculated and were considered as significantly enriched below a cutoff of 0.05.

Sequenced DNA samples were analysed as described previously (Reichelt et al. 2016).

4.4 Identification of SmAP1 co-purified proteins via MS analysis

P. furiosus cells were lysed as described above. Before purification total protein concentrations of the cell extracts were quantified using the QubitTM Protein Assay Kit 3.0 Fluometer (ThermoFisher). For further analysis, 34 mg total protein in 2 ml total volume were used per sample: Strep-SmAP1 (untreated), Strep-SmAP1 (DNase/RNase treated) and Strep-Tag (untreated). The DNase/RNase treated samples were digested using 2 µl Benzonase® Nuclease (Sigma-Aldrich; ≥250 U/µl) and 200 µl RNase Cocktail™ enzyme mix (Invitrogen). All samples were incubated at 37 °C for 60 min and the cell extracts were loaded onto a gravity column containing 250 µl Strep-Tactin® Sepharose® resin (iba). After 10 washing steps (1 ml) using Strep100 buffer, elution was performed using a 1× Elution buffer E (IBA Lifesciences). The following consecutive elution volumes were used: 200 µl, 350 µl, 100 µl. Success of purification was analysed by SDS-PAGE. Three independent experiments were performed per sample.

For mass spectrometry analysis 45 µl of elution step 2 were mixed with NuPAGE LDS sample buffer (4×) (ThermoFisher) and separated using a gradient gel. A gel lane was cut into 10 consecutive slices, which were then transferred into 2 ml micro tubes (Eppendorf) and washed with 50 mM NH_4HCO_3, 50 mM NH_4HCO_3/acetonitrile (3/1) and 50 mM NH_4HCO_3/acetonitrile (1/1) while shaking gently in an orbital shaker (VXR basic Vibrax, IKA). Gel pieces were lyophilized after shrinking by 100 % acetonitrile. To block cysteines, reduction with DTT was carried out for 30 min at 57 °C followed by an alkylation step with iodoacetamide for 30 min at room temperature in the dark. Subsequently, gel slices were washed and lyophilized again as described above. Proteins were

subjected to *in gel* tryptic digest overnight at 37 °C with approximately 2 µg trypsin per 100 µl gel volume (Trypsin Gold, mass spectrometry grade, Promega). Peptides were eluted twice with 100 mM NH_4HCO_3 followed by an additional extraction with 50 mM NH_4HCO_3 in 50 % acetonitrile. Prior to LC-MS/MS analysis, combined eluates were lyophilized and reconstituted in 20 µl of 1 % formic acid. Separation of peptides by reversed-phase chromatography was carried out on an UltiMate 3000 RSLCnano System (Thermo Scientific, Dreieich) which was equipped with a C18 Acclaim Pepmap100 preconcentration column (100 µm i.d. × 20 mm, Thermo Fisher) in front of an Acclaim Pepmap100 C18 nano column (75 µm i.d. × 150 mm, Thermo Fisher). A linear gradient of 4 %–40 % acetonitrile in 0.1 % formic acid over 90 min was used to separate peptides at a flow rate of 300 nl/min. The LC-system was coupled on-line to a maXis plus UHR-QTOF System (Bruker Daltonics, Bremen) via a CaptiveSpray nanoflow electrospray source (Bruker Daltonics). Data-dependent acquisition of MS/MS spectra by CID fragmentation was performed at a resolution of minimum 60000 for MS and MS/MS scans. The MS spectra rate of the precursor scan was 2 Hz processing a mass range between *m*/*z* 175 and *m*/*z* 2000. Via the Compass 1.7 acquisition and processing software (Bruker Daltonics) a dynamic method with a fixed cycle time of 3 s and an *m*/*z* dependent collision energy adjustment between 34 and 55 eV was applied. Raw data processing was performed in Data Analysis 4.2 (Bruker Daltonics), and Protein Scape 3.1.3 (Bruker Daltonics) in connection with Mascot 2.5.1 (Matrix Science) facilitated database searching of the Uniprot *P. furiosus* database (Entry: UP000324354). Search parameters were as follows: enzyme specificity trypsin with 1 missed cleavage allowed, precursor tolerance 0.02 Da, MS/MS tolerance 0.04 Da, carbamidomethylation or propionamide modification of cysteine, oxidation of methionine, deamidation of asparagine and glutamine were set as variable modifications. Mascot peptide ion-score cut-off was set to 25. Search conditions were adjusted to provide a false discovery rate of less than 1 %. Protein list compilation was done using the Protein Extractor function of Protein Scape.

Identification of specific interactions was done as described by Pluchon et al. (2013). Identification of a protein was significant when, at least, five independent peptides covering at least 10 % of the protein sequence were identified and when the resulting mascot score (*M*) was over 100. A noise/signal significance ratio (SR) was defined for each identified protein as described Pluchon et al. Treated and untreated samples were analysed separately, and only interactions are shown, which were identified in all three replicates. The resulting interaction networks were visualized using cytoscape software (Saito et al. 2012). Moreover, enrichment analysis of arCOGs were performed as described above using a p-value cutoff of <0.05.

4.5 Western blot analysis

Western blots were performed as described previously (Waege et al. 2010). The mouse monoclonal antibody that recognizes the *Strep*-tag II epitope was obtained from QIAGEN.

4.6 Heterologous overexpression of SmAP1 in *E. coli* and affinity purification

The Pfu *smap1* gene was amplified using the primers 1542GApproexfor and PF1542Strepgapproexrev, which in addition to the amplification of the *smap1* gene allowed direct fusion of the StrepII-tag sequence to the C-terminus. This PCR fragment was cloned via Gibson assembly (Gibson et al. 2009) into the vector pPROEX HTb (Invitrogen), which was PCR-amplified using the primers pproexgarev and pproextgagafor2. This resulted in a plasmid backbone without the sequence for a 6xHis-Tag and TEV-site. Correctness of the final expression plasmid was assessed by Sanger sequencing (Microsynth). The plasmid was transformed into *E. coli* JW4130*Δhfq* (http://cgsc.biology.yale.edu/KeioList.php) for protein expression. 1 L LB medium containing 25 µg/mL kanamycin and 100 µg/mL ampicillin were inoculated to reach a final OD600 of 0.1–0.2. Cultures were grown at 37 °C to an OD of 0.4–0.6. The culture was cooled to 20 °C and expression of PfuSmAP1 was induced by adding 0.5 mM IPTG. After growth for 10 h at 20 °C, the cells were harvested and the cell pellet stored at −20 °C. A cell pellet corresponding to 200 ml overexpression culture was resuspended in 2 ml Strep100 buffer supplemented with 1× cOmplete™ protease inhibitor cocktail (Roche), lysozyme was added and incubated for 30–60 min on ice. The cells were lysed for 6 × 3 min by sonication using a Bandelin Electronic™ Sonopuls™ HD 2070 homogenizer (cycle: 60 %, power 60 %) in presence of >250 U Pierce Universal Nuclease for Cell Lysis (ThermoFisher) and centrifuged at 15,000 *g* for 30 min at 4 °C. The lysate (700 µl) was treated with 1 µl Pierce Universal Nuclease for Cell Lysis (>250 U/µl) (ThermoFisher) and 25 µl RNase Cocktail™ enzyme mix (Invitrogen) at 37 °C for 60 min, 64 °C for 20 min and afterwards centrifuged at max. speed for >10 min at 4 °C. The supernatant was incubated with 50 µl MagStrep "type3" XT beads (IBA Lifesciences) overnight at 4 °C. Beads were immobilized using a magnetic rack and consecutively washed (1 ml) 5× with Strep100 buffer followed by five washing steps with Strep2000 buffer (10 mM Tris/HCl pH 8.0, 2000 mM NaCl, 10 % (v/v) glycerol) and again, 5× washing steps with Strep100 buffer. PfuSmAP1 was eluted with 50 µL 1× BXT elution buffer (IBA Lifesciences) at 4 °C for 15 min and stored at −20 °C. Protein concentrations were quantified using the QubitTM Protein Assay Kit 3.0 Fluometer (ThermoFisher).

4.7 Mass photometry

Mass photometer (MP) experiments were performed on a TwoMP instrument (Refeyn, UK) at room temperature. Small wells were created by placing a sealing foil (Grace Bio-Labs, USA) on a microscope slide. Calibration with the standard proteins IgG, TG and BSA was performed prior to sample measurement. Ten µl of Strep100 buffer were pipetted into a well and the focus was adjusted by droplet dilution. Two µl of sample were added to the buffer droplet to adjust the volume to 12 µl, resulting in a final concentration of 8 nM PfuSmAP1 monomer. Measurements and data analysis were performed using the manufacturer's recommended default settings.

4.8 Electro Mobility Shift Assays (EMSAs)

Assays were carried out using 10 nM 5′ Hexachloro-fluorescein (HEX) – labelled 38 nt RNA fragments (U-rich: 5′GAACUUCCAUUUUUUGAAUUUGAGUUUAAUGAACUUUG3′; C-rich: 5′GAACUUCCAUCUCUCGAACUCGAGCUCAAUGAACCUUC3′), increasing amounts of Pfu SmAP heptamer (nM): 0, 1.25, 2.5, 5, 10, 20, 25, 50, 100, 200, 400) in binding buffer (20 mM HEPES–KOH pH 7.4, 100 mM NaCl, 1 mM MgCl2, 0.1 % (v/v) Tween20)

supplemented with 40 mM EDTA and 10 % (v/v) glycerol and incubated at 70 °C for 10 min. Samples were separated by electrophoresis (150 mV, 40 min) using a Tris/Boric Acid/EDTA (TBE) buffer system and a native 6 % TBE gel.

4.9 *In vitro* transcription assays

For promoter-dependent cell-free transcription assays, the promoter region of the *hpyA1* gene including the first transcribed 25 nucleotides (Spitalny and Thomm 2008) was fused with a part of the transcribed region of gene PFDSM3638_05070 harbouring a PfuSmAP1 binding region. This DNA construct was produced by gene synthesis (Thermo Fisher) and cloned into the GeneArt standard vector pMA resulting pMUR652.

Sequence of the *hpyA1* promoter – PFDSM3638_05070 (partial) fusion construct:

*5′*GAATTCGCTCTGAATCCGAAAAGTTTATATATCTCTTTTTTCAAAA AACAAAATGGAAATGTGTTATAAATAAAAGGTTCACAGGGAAATCTTATA GAAGAGGCCATGAAACTTTTCCGTCCCTGGGTTCTCAGAGGTATTGAAGCA CTAACAGAGGATAAGAAAATTCCAATAGATGAACTTCCATTTTTTGAATTT GAGTTTAATGAACTTTGGAGCTCTTACATTAGAACAAAGGATCCATATGTT GTAGGAACTTTAATACACCTAGAGAGCTTCTTTAGTTATGTGAAAAACATT CTTCGTTCCGCAGTTTATTATTTTCTCGGCTCTAGGGG*3′*

The DNA template for transcription reactions was PCR amplified using the primers M13F-bio and M13 rev (−29). Promoter-dependent run-off transcription assays were carried out as described previously (Dexl et al. 2018; Hethke et al. 1999). The transcription buffer (40 mM HEPES–KOH pH 7.3, 250 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA) was supplemented with 0.25 µg/ml BSA, 440 µM ATP, 440 µM GTP, 440 µM CTP, 2.7 µM UTP, 0.049 MBq [α-^{32}P]-UTP (111 TBq/mmol) and with 8.5 nM template DNA, 10.5 nM RNAP, 85 nM TBP, 52 nM TFB and varying amounts of PfuSmAP1 heptamer (0–400 nM). The transcription reactions were incubated at 80 °C for 10 min. Samples were taken after 5 and 10 min. Radiolabeled RNA products were extracted via phenol-chloroform extraction and separated on a 7 M urea/15 % acrylamide gel. The gel was transferred and fixed to a Whatman chromatography paper and scanned using a Typhoon *FLA 7500* imaging system (GE Healthcare Life Sciences). Assays were repeated in three individually performed experiments. Run-off signals were quantified using Image Lab 6.0 software (Bio-Rad).

Acknowledgements: Work in the Grohmann lab was supported by the Deutsche Forschungsgemeinschaft (SFB960-TP7 to D.G.). We would like to thank Dr. Kevin Kramm for support with mass photometry measurements and Dipl. Ing. (FH) Thomas Hader and Simon Dechant for technical assistance to cultivate *P. furiosus* in large-scale bio-fermentors. We thank the Core Unit SysMed at the University of Würzburg for excellent technical support and RNA-seq data generation. This work was supported by the IZKF at the University of Würzburg (project Z-6). Work in the protein mass spectrometry lab was supported by the Deutsche Forschungsgemeinschaft (SFB960-TPZ1 to A.B.) We would like to thank E. Hochmuth for excellent technical support with mass spectrometric measurements.

Research ethics: Not applicable.

Author contributions: RR constructed the P. furiosus mutant strains, performed the co-purification experiments and conducted the cell-free transcription assays. RR performed bioinformatic analysis of the data, with contributions from FG. RR conducted mass spectrometry analyses with contributions from AB. TR performed the gel shift assays and mass photometry measurements. RR, SW, WH and DG supervised this work. DG acquired funding to conduct the work. RR prepared the figures, with contribution of all authors. RR and DG wrote the paper, with all authors contributing to manuscript revision, and approval of the submitted version. The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: The authors state no conflict of interest. DG is co-founder of Microbify GmbH. However, there are no commercial interests by the company, or any financial support granted by Microbify GmbH.

Research funding: Work in the Grohmann lab was supported by the Deutsche Forschungsgemeinschaft (SFB960- TP7 to D.G.) and work in the protein mass spectrometry lab was supported by the Deutsche Forschungsgemeinschaft (SFB960-TPZ1 to A.B.). Work in the core Unit SysMed was supported by the IZKF at the University of Wurzburg (project Z-6).

Data availability: RNA-seq and DNA-seq raw data are available at the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena) under study accession number PRJEB62458. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD044174 and 10.6019/PXD044174. Additional raw data can be obtained on request from the corresponding author.

References

Achsel, T., Stark, H., and Lührmann, R. (2001). The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. Proc. Natl. Acad. Sci. USA 98: 3685–3689.

Albrecht, M. and Lengauer, T. (2004). Novel Sm-like proteins with long C-terminal tails and associated methyltransferases. FEBS Lett. 569: 18–26.

Andrade, J.M., Dos Santos, R.F., Chelysheva, I., Ignatova, Z., and Arraiano, C.M. (2018). The RNA-binding protein Hfq is important for ribosome biogenesis and affects translation fidelity. EMBO J. 37: e97631.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME suite. Nucleic Acids Res. 43: W39–W49.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

Boudry, P., Piattelli, E., Drouineau, E., Peltier, J., Boutserin, A., Lejars, M., Hajnsdorf, E., Monot, M., Dupuy, B., Martin-Verstraete, I., et al. (2021). Identification of RNAs bound by Hfq reveals widespread RNA partners and a sporulation regulator in the human pathogen *Clostridioides difficile*. RNA Biol. 18: 1931–1952.

Breuer, R., Gomes-Filho, J.-V., and Randau, L. (2021). Conservation of archaeal C/D box sRNA-guided RNA modifications. Front. Microbiol. 12: 654029.

Butland, G., Peregrín-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 433: 531–537.

Cai, H., Roca, J., Zhao, Y.-F., and Woodson, S.A. (2022). Dynamic refolding of OxyS sRNA by the Hfq RNA chaperone. J. Mol. Biol. 434: 167776.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. 14: 1188–1190.

Dar, D., Prasse, D., Schmitz, R.A., and Sorek, R. (2016). Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. Nat. Microbiol. 1: 16143.

Dexl, S., Reichelt, R., Kraatz, K., Schulz, S., Grohmann, D., Bartlett, M., and Thomm, M. (2018). Displacement of the transcription factor B reader domain during transcription initiation. Nucleic Acids Res. 46: 10066–10081.

Fiala, G. and Stetter, K.O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100 °C. Arch. Microbiol. 145: 56–61.

Fischer, S., Benz, J., Späth, B., Maier, L.-K., Straub, J., Granzow, M., Raabe, M., Urlaub, H., Hoffmann, J., Brutschy, B., et al. (2010). The archaeal Lsm protein binds to small RNAs. J. Biol. Chem. 285: 34429–34438.

Franze de Fernandez, M.T., Eoyang, L., and August, J.T. (1968). Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. Nature 219: 588–590.

Frith, M.C., Saunders, N.F.W., Kobe, B., and Bailey, T.L. (2008). Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput. Biol. 4: e1000071.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6: 343–345.

Grünberger, F., Reichelt, R., Bunk, B., Spröer, C., Overmann, J., Rachel, R., Grohmann, D., and Hausner, W. (2019). Next generation DNA-seq and differential RNA-seq allow Re-annotation of the *Pyrococcus furiosus* DSM 3638 genome and provide insights into archaeal antisense transcription. Front. Microbiol. 10: 1603.

Grünberger, F., Reichelt, R., Waege, I., Ned, V., Bronner, K., Kaljanac, M., Weber, N., El Ahmad, Z., Knauss, L., Madej, M.G., et al. (2020). CopR, a global regulator of transcription to maintain copper homeostasis in *Pyrococcus furiosus*. Front. Microbiol. 11: 613532.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38: 576–589.

Hethke, C., Bergerat, A., Hausner, W., Forterre, P., and Thomm, M. (1999). Cell-free transcription at 95 degrees: thermostability of transcriptional components and DNA topology requirements of *Pyrococcus* transcription. Genetics 152: 1325–1333.

Hirtreiter, A., Damsma, G.E., Cheung, A.C.M., Klose, D., Grohmann, D., Vojnic, E., Martin, A.C.R., Cramer, P., and Werner, F. (2010). Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. Nucleic Acids Res. 38: 4040–4051.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 415: 180–183.

Hou, L., Klug, G., and Evguenieva-Hackenberg, E. (2014). Archaeal DnaG contains a conserved N-terminal RNA-binding domain and enables tailing of rRNA by the exosome. Nucleic Acids Res. 42: 12691–12706.

Kambach, C., Walke, S., Young, R., Avis, J.M., La Fortelle, E.d., Raker, V.A., Lührmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell 96: 375–387.

Kambara, T.K., Ramsey, K.M., and Dove, S.L. (2018). Pervasive targeting of nascent transcripts by Hfq. Cell Rep. 23: 1543–1552.

Kilic, T., Thore, S., and Suck, D. (2005). Crystal structure of an archaeal Sm protein from *Sulfolobus solfataricus*. Proteins 61: 689–693.

Knüppel, R., Trahan, C., Kern, M., Wagner, A., Grünberger, F., Hausner, W., Quax, T.E.F., Albers, S.-V., Oeffinger, M., and Ferreira-Cerca, S. (2021). Insights into synthesis and function of KsgA/Dim1-dependent rRNA modifications in archaea. Nucleic Acids Res. 49: 1662–1687.

Kreuzer, M., Schmutzler, K., Waege, I., Thomm, M., and Hausner, W. (2013). Genetic engineering of *Pyrococcus furiosus* to use chitin as a carbon source. BMC Biotechnol. 13: 9.

Krogan, N.J., Peng, W.-T., Cagney, G., Robinson, M.D., Haw, R., Zhong, G., Guo, X., Zhang, X., Canadien, V., Richards, D.P., et al. (2004). High-definition macromolecular composition of yeast RNA-processing complexes. Mol. Cell 13: 225–239.

Kufel, J., Allmang, C., Verdone, L., Beggs, J.D., and Tollervey, D. (2002). Lsm proteins are required for normal processing of pre-tRNAs and their efficient association with La-homologous protein Lhp1p. Mol. Cell. Biol. 22: 5248–5256.

Kusebauch, U., Zaramela, L.S., Wu, W.-J., Almeida, J.P.P.d., Turkarslan, S., de Lomana, L.G.A., Gomes-Filho, J.V., Vêncio, R.Z.N., Moritz, R.L., et al. (2023). A genome-scale atlas reveals complex interplay of transcription and translation in an archaeon. mSystems 8: e0081622.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Li, J., Yue, L., Li, Z., Zhang, W., Zhang, B., Zhao, F., and Dong, X. (2021). aCPSF1 cooperates with terminator U-tract to dictate archaeal transcription termination efficacy. eLife 10, https://doi.org/10.7554/elife.70464.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.

Maier, L.-K., Benz, J., Fischer, S., Alstetter, M., Jaschinski, K., Hilker, R., Becker, A., Allers, T., Soppa, J., and Marchfelder, A. (2015). Deletion of the Sm1 encoding motif in the lsm gene results in distinct changes in the transcriptome and enhanced swarming activity of *Haloferax* cells. Biochimie 117: 129–137.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18: 67–83.

Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between thermococcales, methanococcales, and methanobacteriales. Life (Basel) 5: 818–840.

Małecka, E.M. and Woodson, S.A. (2021). Stepwise sRNA targeting of structured bacterial mRNAs leads to abortive annealing. Mol. Cell 81: 1988–1999.e4.

Märtens, B., Bezerra, G.A., Kreuter, M.J., Grishkovskaya, I., Manica, A., Arkhipova, V., Djinovic-Carugo, K., and Bläsi, U. (2015). The heptameric SmAP1 and SmAP2 proteins of the crenarchaeon *Sulfolobus solfataricus* bind to common and distinct RNA targets. Life (Basel) 5: 1264–1281.

Märtens, B., Hou, L., Amman, F., Wolfinger, M.T., Evguenieva-Hackenberg, E., and Bläsi, U. (2017a). The SmAP1/2 proteins of the crenarchaeon *Sulfolobus solfataricus* interact with the exosome and stimulate A-rich tailing of transcripts. Nucleic Acids Res. 45: 7938–7949.

Märtens, B., Sharma, K., Urlaub, H., and Bläsi, U. (2017b). The SmAP2 RNA binding motif in the 3'UTR affects mRNA stability in the crenarchaeum *Sulfolobus solfataricus*. Nucleic Acids Res. 45: 8957–8967.

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. 8: 209–220.

Menendez-Gil, P. and Toledo-Arana, A. (2020). Bacterial 3'UTRs: a useful resource in post-transcriptional regulation. Front. Mol. Biosci. 7: 617633.

Mikulecky, P.J., Kaw, M.K., Brescia, C.C., Takach, J.C., Sledjeski, D.D., and Feig, A.L. (2004). *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. Nat. Struct. Mol. Biol. 11: 1206–1214.

Møller, T., Franch, T., Højrup, P., Keene, D.R., Bächinger, H.P., Brennan, R.G., and Valentin-Hansen, P. (2002). Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol. Cell 9: 23–30.

Mura, C., Cascio, D., Sawaya, M.R., and Eisenberg, D.S. (2001). The crystal structure of a heptameric archaeal Sm protein: implications for the eukaryotic snRNP core. Proc. Natl. Acad. Sci. USA 98: 5532–5537.

Mura, C., Kozhukhovsky, A., Gingery, M., Phillips, M., and Eisenberg, D. (2003a). The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs). Protein Sci. 12: 832–847.

Mura, C., Phillips, M., Kozhukhovsky, A., and Eisenberg, D. (2003b). Structure and assembly of an augmented Sm-like archaeal protein 14-mer. Proc. Natl. Acad. Sci. USA 100: 4539–4544.

Mura, C., Randolph, P.S., Patterson, J., and Cozen, A.E. (2013). Archaeal and eukaryotic homologs of Hfq: a structural and evolutionary perspective on Sm function. RNA Biol. 10: 636–651.

Nielsen, J.S., Bøggild, A., Andersen, C.B.F., Nielsen, G., Boysen, A., Brodersen, D.E., and Valentin-Hansen, P. (2007). An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from *Methanococcus jannaschii*. RNA 13: 2213–2223.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14: 417–419.

Phung, D.K., Etienne, C., Batista, M., Langendijk-Genevaux, P., Moalic, Y., Laurent, S., Liuu, S., Morales, V., Jebbar, M., Fichant, G., et al. (2020). RNA processing machineries in Archaea: the 5'-3' exoribonuclease aRNase J of the β-CASP family is engaged specifically with the helicase ASH-Ski2 and the 3'-5' exoribonucleolytic RNA exosome machinery. Nucleic Acids Res. 48: 3832–3847.

Pluchon, P.-F., Fouqueau, T., Crezé, C., Laurent, S., Briffotaux, J., Hogrel, G., Palud, A., Henneke, G., Godfroy, A., Hausner, W., et al. (2013). An extended network of genomic maintenance in the archaeon *Pyrococcus abyssi* highlights unexpected associations between eucaryotic homologs. PloS One 8: e79707.

Portnoy, V. and Schuster, G. (2006). RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. Nucleic Acids Res. 34: 5923–5931.

Reichelt, R., Gindner, A., Thomm, M., and Hausner, W. (2016). Genome-wide binding analysis of the transcriptional regulator TrmBL1 in *Pyrococcus furiosus*. BMC Genom. 17: 40.

Reichelt, R., Grohmann, D., and Willkomm, S. (2018). A journey through the evolutionary diversification of archaeal Lsm and Hfq proteins. Emerg. Top. Life Sci. 2: 647–657.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29: 24–26.

Rodgers, M.L., O'Brien, B., and Woodson, S.A. (2023). Small RNAs and Hfq capture unfolded RNA target sites during transcription. Mol. Cell 83: 1489e.5–1501.e5.

Ross, J.A., Ellis, M.J., Hossain, S., and Haniford, D.B. (2013). Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. RNA 19: 670–684.

Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., Lotia, S., Pico, A.R., Bader, G.D., and Ideker, T. (2012). A travel guide to Cytoscape plugins. Nat. Methods 9: 1069–1076.

Sanders, T.J., Wenck, B.R., Selan, J.N., Barker, M.P., Trimmer, S.A., Walker, J.E., and Santangelo, T.J. (2020). FttA is a CPSF73 homologue that terminates transcription in Archaea. Nat. Microbiol. 5: 545–553.

Santangelo, T.J. and Reeve, J.N. (2006). Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. J. Mol. Biol. 355: 196–210.

Schumacher, M.A., Pearson, R.F., Møller, T., Valentin-Hansen, P., and Brennan, R.G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J. 21: 3546–3556.

Sedlyarova, N., Shamovsky, I., Bharati, B.K., Epshtein, V., Chen, J., Gottesman, S., Schroeder, R., and Nudler, E. (2016). sRNA-mediated control of transcription termination in *E. coli*. Cell 167: 111.e13–121.e13.

Sledjeski, D.D., Whitman, C., and Zhang, A. (2001). Hfq is necessary for regulation by the untranslated RNA DsrA. J. Bacteriol. 183: 1997–2005.

Spitalny, P. and Thomm, M. (2008). A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. Mol. Microbiol. 67: 958–970.

Sukhodolets, M.V. and Garges, S. (2003). Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. Biochemistry 42: 8022–8034.

Tharun, S., He, W., Mayes, A.E., Lennertz, P., Beggs, J.D., and Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. Nature 404: 515–518.

Thore, S., Mayer, C., Sauter, C., Weeks, S., and Suck, D. (2003). Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. J. Biol. Chem. 278: 1239–1247.

Törö, I., Basquin, J., Teo-Dreher, H., and Suck, D. (2002). Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus*. J. Mol. Biol. 320: 129–142.

Törö, I., Thore, S., Mayer, C., Basquin, J., Séraphin, B., and Suck, D. (2001). RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. EMBO J. 20: 2293–2303.

Trouillon, J., Han, K., Attrée, I., and Lory, S. (2022). The core and accessory Hfq interactomes across *Pseudomonas aeruginosa* lineages. Nat. Commun. 13: 1258.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 403: 623–627.

UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res 51: D523–D531.

Valentin-Hansen, P., Eriksen, M., and Udesen, C. (2004). The bacterial Sm-like protein Hfq: a key player in RNA transactions. Mol. Microbiol. 51: 1525–1533.

van Assche, E., van Puyvelde, S., Vanderleyden, J., and Steenackers, H.P. (2015). RNA-binding proteins involved in post-transcriptional regulation in bacteria. Front. Microbiol. 6: 141.

Vogel, J. and Luisi, B.F. (2011). Hfq and its constellation of RNA. Nat. Rev. Microbiol. 9: 578–589.

Waege, I., Schmid, G., Thumann, S., Thomm, M., and Hausner, W. (2010). Shuttle vector-based transformation system for Pyrococcus furiosus. Appl. Environ. Microbiol. 76: 3308–3313.

Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev. 15: 1637–1651.

Weixlbaumer, A., Grünberger, F., Werner, F., and Grohmann, D. (2021). Coupling of transcription and translation in archaea: cues from the bacterial world. Front. Microbiol. 12: 661827.

Wilusz, C.J. and Wilusz, J. (2005). Eukaryotic Lsm proteins: lessons from bacteria. Nat. Struct. Mol. Biol. 12: 1031–1036.

Witharana, C., Roppelt, V., Lochnit, G., Klug, G., and Evguenieva-Hackenberg, E. (2012). Heterogeneous complexes of the RNA exosome in *Sulfolobus solfataricus*. Biochimie 94: 1578–1587.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 11: R14.

Yue, L., Li, J., Zhang, B., Qi, L., Li, Z., Zhao, F., Li, L., Zheng, X., and Dong, X. (2020). The conserved ribonuclease aCPSF1 triggers genome-wide transcription termination of Archaea via a 3'-end cleavage mode. Nucleic Acids Res. 48: 9589–9605.

Zhan, X., Yan, C., Zhang, X., Lei, J., and Shi, Y. (2018). Structures of the human pre-catalytic spliceosome and its precursor spliceosome. Cell Res. 28: 1129–1140.

Zhang, A., Wassarman, K.M., Ortega, J., Steven, A.C., and Storz, G. (2002). The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol. Cell 9: 11–22.

Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G., and Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. Mol. Microbiol. 50: 1111–1124.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/hsz-2023-0215).